EXHIBIT 6

"

Guideline for Industry

Detection of Toxicity to Reproduction for Medicinal Products

ICH-S5A

September 1994

TABLE OF CONTENTS

GUIDELINE FOR INDUSTRY[†]

DETECTION OF TOXICITY TO REPRODUCTION FOR MEDICINAL PRODUCTS[††]

I.   INTRODUCTION (1)

A.   Purpose of the Guideline (1.1)

There is a considerable overlap in the methodology that could be used to test chemicals and medicinal products for potential reproductive toxicity. As a first step to using this wider methodology for efficient testing, this guideline attempts to consolidate a strategy based on study designs currently in use for testing of medicinal products; it

---

[†]This guideline was developed within the Expert Working Group (Safety) of the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) and has been subject to consultation by the regulatory parties, in accordance with the ICH process. This document has been endorsed by the ICH Steering Committee at *Step 4* of the ICH process, June 24, 1993. At *Step 4* of the process, the final draft is recommended for adoption to the regulatory bodies of the European Union, Japan and the USA. This guideline was published in the Federal Register on September 22, 1994 (59 FR 48746) and is applicable to drug and biological products. In the past, guidelines have generally been issued under § 10.90(b) [21 CFR 10.90(b)], which provides for the use of guidelines to state procedures or standards of general applicability that are not legal requirements but that are acceptable to FDA. The agency is now in the process of revising §10.90(b). Therefore, this guideline is not being issued under the authority of §10.90(b), and it does not create or confer any rights, privileges or benefits for or on any person, nor does it operate to bind FDA in any way. For additional copies of this guideline contact the Executive Secretariat Staff, HFD-8, Center for Drug Evaluation and Research, 7500 Standish Place, Rockville, MD, 20855, 301-594-1012. An electronic version of this guideline is also available via Internet by connecting to the CDER FTP server (CDVS2.CDER.FDA.GOV) using the FTP protocol.

[††]To help facilitate understanding of the guideline, the agency is providing further clarification of important questions that have been raised since initial general distribution of the document at ICH 2 by both industry and regulatory scientists. See Appendix B for these clarifications.

should encourage the full assessment on the safety of chemicals on the development of the offspring. It is perceived that tests in which animals are treated during defined stages of reproduction better reflect human exposure to medicinal products and allow more specific identification of stages at risk. While this approach may be useful for most medicines, long-term exposure to low doses does occur and may be represented better by a one- or two-generation study approach. The actual testing strategy should be determined by:

- Anticipated drug use especially in relation to reproduction,

- The form of the substance and route(s) of administration intended for humans, and

- Making use of any existing data on toxicity, pharmacodynamics, kinetics, and similarity to other compounds in structure/activity.

To employ this concept successfully, flexibility is needed.[1] No guideline can provide sufficient information to cover all possible cases. All persons involved should be willing to discuss and consider variations in test strategy according to the state-of-the-art and ethical standards in human and animal experimentation. Areas where more basic research would be useful for optimization of test designs are male fertility assessment, and kinetic and metabolism in pregnant/lactating animals.

B.   Aim of Studies (1.2)

The aim of reproduction toxicity studies is to reveal any effect of one or more active substance(s) on mammalian reproduction. For this purpose, both the investigations and the interpretation of the results should be related to all other pharmacological and toxicological data available to determine whether potential reproductive risks to humans are greater, lesser, or equal to those posed by other toxicological manifestations. Further, repeated dose toxicity studies can provide important information regarding potential effects on reproduction, particularly male fertility. To extrapolate the results to humans (assess the relevance), data on likely human exposures, comparative kinetics, and mechanisms of reproductive toxicity may be helpful. The combination of studies selected should allow exposure of mature

adults and all stages of development from conception to sexual maturity. To allow detection of immediate and latent effects of exposure, observations should be continued through one complete life cycle, i.e., from conception in one generation through conception in the following generation. For convenience of testing this integrated sequence can be subdivided into the following stages.

1(A). Premating to conception (adult male and female reproductive functions, development and maturation of gametes, mating behavior, fertilization).

2(B). Conception to implantation (adult female reproductive functions, preimplantation development, implantation).

3(C). Implantation to closure of the hard palate (adult female reproductive functions, embryonic development, major organ formation).

4(D). Closure of the hard palate to the end of pregnancy (adult female reproductive functions, fetal development and growth, organ development and growth).

5(E). Birth to weaning (adult female reproductive functions, neonate adaption to extrauterine life, preweaning development and growth).

6(F). Weaning to sexual maturity (postweaning development and growth, adaption to independent life, attainment of full sexual function).

For timing conventions see Note 2.[2]

C.   Choice of Studies (1.3)

The guideline addresses the design of studies primarily for detection of effects on reproduction. When an effect is detected, further studies to characterize fully the nature of the response have to be designed on a case-by-case basis.[3] The rationale for the set of studies chosen should be given and should include an explanation for the choice of dosages.

3

Studies should be planned according to the "state-of-the art," and take into account preexisting knowledge of class-related effects on reproduction. They should avoid suffering and should use the minimum number of animals necessary to achieve the overall objectives. If a preliminary study is performed, the results should be considered and discussed in the overall evaluation.[4]

II.   ANIMAL CRITERIA (2)

The animals used should be well defined with respect to their health, fertility, fecundity, prevalence of abnormalities, embryofetal deaths, and the consistency they display from study to study. Within and between studies, animals should be of comparable age, weight, and parity at the start; the easiest way to fulfill these criteria is to use animals that are young, mature adults at the time of mating with the females being virgin.

A.   Selection and Number of Species (2.1)

Studies should be conducted in mammalian species. It is generally desirable to use the same species and strain as in other toxicological studies. Reasons for using rats as the predominant rodent species are practicality, comparability with other results obtained in this species and the large amount of background knowledge accumulated.

In embryotoxicity studies only, a second mammalian species traditionally has been required, the rabbit being the preferred choice as a "nonrodent." Reasons for using rabbits in embryotoxicity studies include the extensive background knowledge that has accumulated, as well as availability and practicality. Where the rabbit is unsuitable, an alternative nonrodent or a second rodent species may be acceptable and should be considered on a case-by-case basis.[5]

B.   Other Test Systems (2.2)

Other test systems are considered to be any developing mammalian and nonmammalian cell systems, tissues, organs, or organism cultures developing independently in vitro or in vivo. Integrated with whole animal studies either for priority selection within homologous series or as secondary investigations to elucidate mechanisms of action, these systems can provide invaluable information and,

4

indirectly, reduce the numbers of animals used in experimentation. However, they lack the complexity of the developmental processes and the dynamic interchange between the maternal and the developing organisms. These systems cannot provide assurance of the absence of effect nor provide perspective in respect of risk/exposure. In short, there are no alternative test systems to whole animals currently available for reproduction toxicity testing with the aims set out in the introduction.[6]

III.   GENERAL RECOMMENDATIONS CONCERNING TREATMENT (3)

A.   Dosages (3.1)

Selection of dosages is one of the most critical issues in design of the reproductive toxicity study. The choice of the high dose should be based on data from all available studies (pharmacology, acute and chronic toxicity and kinetic studies[7]). A repeated dose toxicity study of about 2 to 4 weeks duration provides a close approximation to the duration of treatment in segmental designs of reproductive studies. When sufficient information is not available, preliminary studies are advisable.[4]

Having determined the high dosage, lower dosages should be selected in a descending sequence, the intervals depending on kinetic and other toxicity factors. Whilst it is desirable to be able to determine a "no observed adverse effect level," priority should be given to setting dosage intervals close enough to reveal any dosage-related trends that may be present.[8]

B.   Route and Frequency of Administration (3.2)

In general the route or routes of administration should be similar to those intended for human usage. One route of substance administration may be acceptable if it can be shown that a similar distribution (kinetic profile) results from different routes.[9]

The usual frequency of administration is once daily but consideration should be given to use either more frequent or less frequent administration taking kinetic variables into account.[10]

C.   Kinetics (3.3)

It is preferable to have some information on kinetics before initiating reproduction studies since this may suggest the need to adjust choice of species, study design, and dosing schedules. At this time, the information need not be sophisticated nor derived from pregnant or lactating animals.

At the time of study evaluation, further information on kinetics in pregnant or lactating animals may be required according to the results obtained.[10]

D.   Control Groups (3.4)

It is recommended that control animals be dosed with the vehicle at the same rate as test group animals. When the vehicle may cause effects or affect the action of the test substance, a second (sham- or untreated) control group should be considered.

IV.   PROPOSED STUDY DESIGNS-COMBINATION OF STUDIES (4)

All available pharmacological, kinetic, and toxicological data for the test compound and similar substances should be considered in deciding the most appropriate strategy and choice of study design. It is anticipated that, initially, preference will be given to designs that do not differ too radically from those of established guidelines for medicinal products (the most probable option). For most medicinal products, the three-study design will usually be adequate. Other strategies, combinations of studies, and study designs could be as valid or more valid as the "most probable option" according to circumstances. The key factor is that, in total, they leave no gaps between stages and allow direct or indirect evaluation of all stages of the reproductive process.[11]

Designs should be justified.

A.   The Most Probable Option (4.1)

The most probable option can be equated to a combination of studies for effects on:

- Fertility and early embryonic development,

- Prenatal and postnatal development, including maternal function, and

- Embryo-fetal development.

1.   Study of Fertility and Early Embryonic Development to Implantation (4.1.1)

   a.   Aim

   To test for toxic effects/disturbances resulting from treatment from before mating (males/females) through mating and implantation. This comprises evaluation of stages A and B of the reproductive process [see I.B. (1.2)]. For females this should detect effects on the oestrous cycle, tubal transport, implantation, and development of preimplantation stages of the embryo. For males it will permit detection of functional effects (e.g., on libido, epididymal sperm maturation) that may not be detected by histological examinations of the male reproductive organs.[12]

   b.   Assessment of:

   - Maturation of gametes,

   - Mating behavior,

   - Fertility,

   - Preimplantation stages of the embryo, and

   - Implantation.

   c.   Animals

   At least one species, preferably rats.

   d.   Number of Animals

7

The number of animals per sex per group should be sufficient to allow meaningful interpretation of the data.[13]

e.   Administration Period

The design assumes that, especially for effects on spermatogenesis, use will be made of data from repeated dose toxicity studies of at least 1-month duration. Provided no effects have been found that preclude this, a premating treatment interval of 2 weeks for females and 4 weeks for males can be used.[12] Selection of the length of the premating administration period should be stated and justified [see also I.A. (1.1), pointing out the need for research]. Treatment should continue throughout mating to termination of males and at least through implantation for females. This will permit evaluation of functional effects on male fertility that cannot be detected by histologic examination in repeated dose toxicity studies and effects on mating behavior in both sexes. If data from other studies show there are effects on weight or histologic appearance of reproductive organs in males or females, or if the quality of examinations is dubious or if there are no data from other studies, then a more comprehensive study should be designed.[12]

f.   Mating

A mating ratio of 1:1 is advisable and procedures should allow identification of both parents of a litter.[14]

g.   Terminal Sacrifice

Females may be sacrificed at any point after midpregnancy.

Males may be sacrificed at any time after mating but it is advisable to ensure successful induction of pregnancy before taking such an irrevocable step.[15]

h.   Observations

8

During study:

- Signs and mortalities at least once daily;

- Body weight and body weight changes at least twice weekly;[16]

- Food intake at least once weekly (except during mating);

- Record vaginal smears daily, at least during the mating period, to determine whether there are effects on mating or precoital time; and

- Observations that have proved of value in other toxicity studies.

At terminal examination:

- Necropsy (macroscopic examination) of all adults;

- Preserve organs with macroscopic findings for possible histological evaluation; keep corresponding organs of sufficient controls for comparison;

- Preserve testes, epididymides, ovaries and uteri from all animals for possible histological examination and evaluation on a case-by-case basis; tissues can be discarded after completion and reporting of the study;

- Sperm count in epididymides or testes, as well as sperm viability;

- Count corpora lutea, implantation sites;[16] and

- Live and dead conceptuses.

9

2.   Study for Effects on Prenatal and Postnatal Development, Including Maternal Function (4.1.2)

    a.   Aim

        To detect adverse effects on the pregnant/lactating female and on development of the conceptus and the offspring following exposure of the female from implantation through weaning. Since manifestations of effect induced during this period may be delayed, observations should be continued through sexual maturity [i.e., stages 3 to 6 (C to F) listed in I.B. (1.2)].[17,18]

    b.   Adverse Effects To Be Assessed

        •   Enhanced toxicity relative to that in nonpregnant females;

        •   Prenatal and postnatal death of offspring;

        •   Altered growth and development; and

        •   Functional deficits in offspring, including behavior, maturation (puberty), and reproduction (F1).

    c.   Animals

        At least one species, preferably rats.

    d.   Number of Animals

        The number of animals per sex per group should be sufficient to allow meaningful interpretation of the data.[13]

    e.   Administration Period

        Females are exposed to the test substance from implantation to the end of lactation [i.e., stages 3 to 5 (C to E) listed in I.B. (1.2)].

10

f.   Experimental Procedure

The females are allowed to deliver and rear their offspring to weaning at which time one male and one female offspring per litter should be selected (document method used) for rearing to adulthood and mating to assess reproductive competence.[19]

g.   Observations

During study (for maternal animals):

- Signs and mortalities at least once daily,

- Body weight and body weight change at least twice weekly,[16]

- Food intake at least once weekly at least until delivery,

- Observations that have proved of value in other toxicity studies,

- Duration of pregnancy, and

- Parturition.

At terminal examination (for maternal animals and where applicable for offspring):

- Necropsy (macroscopic examination) of all adults;

- Preservation and possibly histological evaluation of organs with macroscopic findings; keep corresponding organs of sufficient controls for comparison;

- Implantations;[16]

- Abnormalities;

11

- Live offspring at birth;

- Dead offspring at birth;

- Body weight at birth;

- Preweaning and postweaning survival and growth/body weight,[20] maturation, and fertility;

- Physical development;[21]

- Sensory functions and reflexes;[21] and

- Behavior.[21]

3. Study for Effects on Embryo-Fetal Development (4.1.3)

   a. Aim

   To detect adverse effects on the pregnant female and development of the embryo and fetus consequent to exposure of the female implantation to closure of the hard palate [i.e., stages 3 to 4 (C to D) listed in I.B. (1.2)].

   b. Adverse Effects To Be Assessed

   - Enhanced toxicity relative to that in nonpregnant females,

   - Embryofetal death,

   - Altered growth, and

   - Structural changes.

   c. Animals

   Usually, two species: one rodent, preferably rats; one

nonrodent, preferably rabbits.[5]   Justification should be provided when using one species.

d.   Number of Animals

The number of animals should be sufficient to allow meaningful interpretation of the data.[13]

e.   Administration Period

The treatment period extends from implantation to the closure of the hard palate [i.e., end of 3 (C), see I.B. (1.2)].

f.   Experimental Procedure

Females should be sacrificed and examined about 1 day prior to parturition. All fetuses should be examined for viability and abnormalities. To allow subsequent assessment of the relationship between observations made by different techniques fetuses should be individually identified.[22]

When using techniques requiring allocation to separate examination for soft tissue or skeletal changes, it is preferable that 50 percent of fetuses from each litter be allocated for skeletal examination. A minimum of 50 percent rat fetuses should be examined for visceral alterations, regardless of the technique used. When using fresh microdissection techniques for soft tissue alterations--which is the strongly preferred method for rabbits--100 percent of rabbit fetuses should be examined for soft tissue and skeletal abnormalities.

g.   Observations

During study (for maternal animals):

- Signs and mortalities at least once daily,

- Body weight and body weight change at

least twice weekly,[16]

- Food intake at least once weekly, and

- Observations that have proved of value in other toxicity studies.

At terminal examination:

- Necropsy (macroscopic examination) of all adults;

- Preserve organs with macroscopic findings for possible histological evaluation; keep corresponding organs of sufficient controls for comparison;

- Count corpora lutea, numbers of live and dead implantations;[16]

- Individual fetal body weight;

- Fetal abnormalities;[22] and

- Gross evaluation of placenta.

B.   Single Study Design (rodents) (4.2)

If the dosing period of the fertility study and prenatal and postnatal study are combined into a single investigation, this comprises evaluation of stages 1 to 6 (A to F) of the reproductive process [see I.B. (1.2)]. If such a study, if it includes fetal examinations, provided clearly negative results at sufficiently high exposure, no further reproduction studies in rodents should be required. Fetal examinations for structural abnormalities can also be supplemented with an embryo-fetal development study (or studies) to make a two-study approach.[3,11]

Results from a study for effects on embryo-fetal development in a second species are expected [see also IV.A.3. (4.1.3)].

C.   Two Study Design (rodents) (4.3)

The simplest two-segment design would consist of the fertility study and the prenatal and postnatal development study, if it includes fetal examinations. It can be assumed, however, that if the prenatal and postnatal development study provided no indication of prenatal effects at adequate margins above human exposure, the additional fetal examinations [see IV.A.3. (4.1.3)] are most unlikely to provide a major change in the assessment of risk.

Alternatively, female treatment in the fertility study [IV.A.1. (4.1.1)] could be continued until closure of the hard palate and fetuses examined according to the procedures of the embryo-fetal development study [IV.A.3. (4.1.3)]. This, combined with the prenatal and postnatal study [IV.A.2. (4.1.2)] would provide all the examinations required in "the most probable option" but use considerably less animals.[3,11]

Results from a study for effects on embryo-fetal development in a second species are expected [see also IV.A.3. (4.1.3)].

V.   STATISTICS (5)

Analysis of the statistics of a study is the means by which results are interpreted. The most important part of this analysis is to establish the relationship between the different variables and their distribution (descriptive statistics), because these determine how groups should be compared. The distributions of the endpoints observed in reproductive tests are usually nonnormal and extend from almost continuous to the extreme categorical. When employing inferential statistics (determination of statistical significance) the mating pair or litter, not the fetus or neonate, should be used as the basic unit of comparison. The tests used should be justified.[23]

VI.   DATA PRESENTATION (6)

The key to good reporting is the tabulation of individual values in a clear concise manner to account for every animal that was entered into the study. A reader should be able to follow the history of any individual animal from initiation to termination and should be able to deduce with ease the contribution that the individual has made to any group summary values. Group summary values should be presented in a form that is biologically

plausible (i.e., avoid false precision) and that reflects the distribution of the variable. Appendices or tabulations of individual values such as bodyweight, food consumption, litter values should be concise and, as far as possible, consist of absolute rather than calculated values; unnecessary duplication should be avoided.

For tabulation of low frequency observations such as clinical signs, autopsy findings, abnormalities, etc., it is advisable to group together the (few) individuals with a positive recording. Especially in the presentation of data on structural changes (fetal abnormalities) the primary listing (tabulation) should clearly identify the litters containing abnormal fetuses, identify the affected fetuses in the litter, and report all the changes observed in the affected fetus. Secondary listings by type of change can be derived from this, if necessary.

VII.   TERMINOLOGY (7)

Besides effects on the reproductive competence of adult animals toxicity to reproduction includes:

Developmental toxicity: Any adverse effect induced prior to attainment of adult life. It includes effects induced or manifested in the embryonic or fetal period and those induced or manifested postnatally.

Embryotoxicity, *fetotoxicity*, *embryo-fetal toxicity*: Any adverse effect on the conceptus resulting from prenatal exposure, including structural or functional abnormalities or postnatal manifestations of such effects. Terms like "embryotoxicity" or "fetotoxicity" relate to the timepoint/-period of induction of adverse effects, irrespective of the time of detection.

One-, two-, or three-generation studies: Are defined according to the number of adult breeding generations directly exposed to the test material. For example, in a one-generation study there is direct exposure of the F0 generation and indirect exposure (via the mother) of the F1 generation, and the study is usually terminated at the weaning of the F1 generation. In a two-generation study as used for agro-chemicals and industrial chemicals there is direct exposure of the F0 generation, indirect and direct exposure of the F1 generation and indirect exposure of the F2 generation. A three-generation study is defined accordingly.

Body burden: The total internal dosage of an individual arising from the

16

administration of a substance, comprising parent compound and metabolites, taking distribution and accumulation into account.

Kinetics: The term "kinetics" is used consistently throughout this guideline, irrespective of intending to mean pharmaco- and/or toxicokinetics. No better single term was available.

VIII. REFERENCE

Federal Register, Vol. 59, No.183, Thursday, September 22, 1994, pages 48746-48752.

APPENDIX B

ADDITIONAL CLARIFICATION FOR USE OF GUIDELINE

To help facilitate understanding of the guideline, the agency is providing further clarification of important questions that have been raised since initial general distribution of the document at ICH 2 by both industry and regulatory scientists.

General Comments

First pass tests in the guideline are those tests that will are likely be performed as general screens (i.e., the three-study design or "most probable option") to identify potential treatment related effects. Secondary tests are those designed to characterize, e.g., the nature, scope, and/or origin of the toxic effect. In general, repeated dose general toxicity studies of 2 to 4 weeks duration may provide a close approximation of the doses to be used in the reproductive toxicology studies.

Male Fertility

As stated in the introduction to the guideline, studies are ongoing to optimize parameters to be used in fertility studies, including the optimal treatment period for males prior to mating, histological techniques for the evaluation of sex organs, and techniques to evaluate sperm. It is expected that, in most cases, viability will be measured indirectly by evaluating sperm motility. A variety of methods will be acceptable to evaluate sperm, including vital dye staining, flow cytometric analysis, and nonautomated and automated methods to measure the percent of motile sperm. Sponsors should justify the methods used and define the objective criteria established to assess the data obtained. It is expected that improvements in methods to assess male reproductive performance will evolve over the next few years.

The design of the study of fertility [IV.A.1. (4.1.1)] assumes that, especially for effects on spermatogenesis, use will be made of data from repeated dose toxicity studies of at least 1-month duration. The agency encourages the use of good pathological and histopathological examination techniques in the repeated dose toxicity studies in addition to the staging of spermatogenesis which is routinely employed. The preservation of testes and epididymides from all animals from ICH study IV.A.1. (4.1.1) provides

an opportunity for more detailed histopathological examination on a case-by-case basis; for example, if unexpected effects on sperm count or viability are observed. There may be cases due to species-specific effects or technical considerations (e.g., multiple samplings are required over time) when sperm evaluation in nonrodents may be more appropriate.

The duration of pretreatment for males in ICH study IV.A.1. (4.1.1) is 4 weeks, unless data from other studies suggest that this should be modified. Males should be treated throughout the mating period (generally between 2 and 3 weeks) and at least through implantation of the females. Thus, males will generally be sacrificed following at least 7 to 9 weeks dosing. Evaluations should generally include organ weights and macroscopic examinations of testis, epididymis, seminal vesicle, and prostate. Sperm counts and sperm viability (e.g., motility) should be assessed. Tissues should be saved for potential histological assessment, as such assessments may be required on a case-by-case basis. If histological data are not available from previous studies or the quality of the data are dubious, then histological evaluation should be performed in this study.

Prenatal and Postnatal Development

When studying the effect on postnatal development, the reduction of litter size by culling is still under discussion. If culling is performed, it should be randomized. Whether or not it is performed, it should be explained by the investigator. Observations on offspring in ICH study IV.A.2. (4.1.2) include sensory functions and reflexes and behavior, consistent with previous guidelines from Japan and the European Union. Specific functional tests have not been recommended in the ICH guideline. Investigators are encouraged to use methods that will assess sensory functions, motor activity, learning, and memory to help characterize functional deficits in offspring. Under the terminology section of the guideline, a three-generation study is defined as direct exposure of the F0 generation, indirect and direct exposure of the F1 and F2, and indirect exposure of the F3 generation.

<div align="right">**APPENDIX A**</div>

<div align="center">**NOTES FOR CLARIFICATION OF GUIDELINE**</div>

1.   Scientific Flexibility [I.A. (1.1)]

These guidelines are not mandatory rules, they are a starting point rather than an endpoint. They provide a basis from which an investigator can devise a strategy for testing according to available knowledge of the test material and the state-of- the art. For encouragement, some alternative test designs have been mentioned in this document but there are others that can be sought out or devised. In devising a strategy, the primary objective should be to detect and bring to light any indication of toxicity to reproduction.

Fine details of study design and technical procedures have been omitted from the text. Such decisions rightly belong in field of the investigator since a technique that may be suitable for one laboratory may not be suitable in another. The investigator needs to utilize staff and resources to do the best he or she can achieve and should know how to do this better than any outsider; human attributes of attitude, ability, and consistency are more important than material facilities. For necessary compliance to good laboratory practices (GLP), reference is made to such regulations.

2.   Timing Conventions [I.B. (1.2)]

In this guideline the convention for timing of pregnancy is to refer to the day that a sperm-positive vaginal smear and/or plug is observed as day 0 of pregnancy even if mating occurs overnight. Unless shown otherwise it is assumed that, for rats, mice and rabbits implantation occurs on day 6-7 of pregnancy, and closure of the hard palate on day 15-18 of pregnancy.

Other conventions are equally acceptable if defined in reports. Also, the investigator should be consistent in different studies to ensure that no gaps in treatment occur. It is an advisable precaution to provide an overlap of at least 1 day in the exposure period of related studies.

The accuracy of the time of mating should be specified because this will affect the variability of fetal and neonatal parameters.

Similarly, for reared litters, the day offspring are born will be considered as

postnatal or lactation day 0 unless otherwise specified. However, particularly with regard to delays in, or prolongation of, parturition, reference to a postcoital timeframe may be useful.

3.   First Pass and Secondary Testing [I.C. (1.3)]

To a greater or lesser degree, all first pass (guideline) tests are apical in nature, i.e., an effect on one endpoint may have several different origins. A reduced litter size at birth may be due to a reduced ovulation rate (corpora lutea count), higher rate of preimplantation deaths, higher rate of postimplantation deaths, or immediate postnatal deaths. In turn, these deaths may be the consequence of an earlier physical malformation that can no longer be observed due to subsequent secondary changes and so on. Particularly for effects with a natural low frequency among controls, discrimination between treatment-induced and coincidental occurrence is dependent upon association with other types of effects.

A toxicant usually induces more than one type of effect in a dose-dependent manner. For example, induction of malformation is almost invariably associated with increased embryonic death and an increased incidence of less severe structural changes. Given an effect on one endpoint, secondary investigations for possible associations should be considered, i.e., the nature, scope, and origins of the substance's toxicity should be characterized. Characterization should also include identification of dose-response relationships to facilitate risk assessment; this is different from the situation in first pass tests where the presence or absence of a dose response assists discrimination between treatment-related and coincidental differences.

4.   Preliminary Studies [I.C. (1.3)]

At the time most reproduction studies are planned or initiated there is usually information available from acute and repeated dose toxicity studies of at least 1-month duration. This information can be expected to be sufficient in identifying doses for reproductive studies. If adequate preliminary studies are performed, they are part of the justification of the choice of dose for the main study. Such studies should be submitted regardless of their GLP-status in principle. This may avoid unnecessary use of animals.

5.   Selection of Species and Strains [II.A. (2.1)]

In choosing an animal species and strain for reproductive toxicity testing, care should be given to select a relevant model. Selection of the species and strain used in other toxicology studies may avoid the need for additional preliminary studies. If it can be shown--by means of kinetic, pharmacological, and toxicological data--that the species selected is a relevant model for the human, a single species can be sufficient. There is little value in using a second species if it does not show the same similarities to humans. Advantages and disadvantages of species (strains) should be considered in relation to the substance to be tested, the selected study design, and in the subsequent interpretation of the results.

All species have their advantages. Rats, and to a lesser extent mice, are good general purpose models; the rabbit has been somewhat neglected as a "nonrodent" species for repeated dose toxicity and other reproduction studies than embryotoxicity testing. It has attributes that would make it a useful model for fertility studies, especially male fertility. For both rabbits and dogs (which are often used as a second species for chronic toxicity studies) it is feasible to obtain semen samples without resorting to painful techniques (electro ejaculation) for longitudinal semen analysis. Most of the other species are not good, general purpose models and probably are best used for very specific investigations only.

All species have their disadvantages, for example:

- *Rats*: Sensitivity to sexual hormones, unsuitable for dopamine agonists due to dependence on prolactin as the primary hormone for establishment and maintenance of early pregnancy, highly susceptible to nonsteroidal anti-inflammatory drugs in late pregnancy.

- *Mice*: Fast metabolic rate, stress sensitivity, malformation clusters (which occur in all species) particularly evident, small fetus.

- *Rabbits*: Often lack of kinetic and toxicity data, susceptibility to some antibiotics and to disturbance of the alimentary tract, clinical signs can be difficult to interpret.

- *Guinea pigs*: Often lack of kinetic and toxicity data, susceptibility to some antibiotics and to disturbance of the alimentary tract, long fetal period, insufficient historical background data.

- *Domestic and/or mini pigs*: Malformation clusters with variable

A-3

background rate, large amounts of compound required, large housing necessary, insufficient historical background data.

- *Ferrets*: Seasonal breeder unless special management systems used (success highly dependent on human/animal interaction), insufficient historical background data.

- *Hamsters*: Intravenous route difficult if not impossible, can hide doses in the cheek pouches and can be very aggressive, sensitive to intestinal disturbance, overly sensitive teratogenic response to many chemicals, small foetus.

- *Dogs*: Seasonal breeders, inbreeding factors, insufficient historical background data.

- *Nonhuman primates*: Kinetically they can differ from humans as much as other species, insufficient historical background data, often numbers too low for detection of risk. They are best used when the objective of the study is to characterize a relatively certain reproductive toxicant, rather than detect a hazard.

6.   Uses of Other Test Systems Than Whole Animals [II.B. (2.2)]

Other tests systems have been developed and used in preliminary investigations ("prescreening" or priority selection) and secondary testing.

For preliminary investigation of a range of analogue series of substances, it is essential that the potential outcome in whole animals is known for at least one member of the series to be studied (by inference, effects are expected). With this strategy, substances can be selected for higher level testing.

For secondary testing or further substance characterization, other test systems offer the possibility to study some of the observable developmental processes in detail, e.g., to reveal specific mechanisms of toxicity, to establish concentration-response relationships, to select 'sensitive periods,' or to detect effects of defined metabolites.

7.   Selection of Dosages [III.A. (3.1)]

Using similar doses in the reproductive toxicity studies as in the repeated dose toxicity studies will allow interpretation of any potential effects on fertility in context with general systemic toxicity.

A-4

Some minimal toxicity is expected to be induced in the high-dose dams.

According to the specific compound, factors limiting the high dosage determined from repeat dose toxicity studies or from preliminary reproduction studies could include:

- Reduction in bodyweight gain;

- Increased bodyweight gain, particularly when related to perturbation of homeostatic mechanisms;

- Specific target organ toxicity;

- Haematology, clinical chemistry;

- Exaggerated pharmacological response, which may or may not be reflected as marked clinical reactions (e.g., sedation, convulsions);

- The physico-chemical properties of the test substance or dosage formulation which, allied to the route of administration, may impose practical limitations in the amount that can be administered; under most circumstances 1 gram per kilogram per day (g/kg/day) should be an adequate limit dose;

- Kinetics can be useful in determining high-dose exposure for low toxicity compounds; there is, however, little point in increasing administered dosage if it does not result in increased plasma or tissue concentration; and

- Marked increase in embryo-fetal lethality in preliminary studies.

8.   Determination of Dose-Response Relationships [III.A. (3.1)]

For many of the variables in reproduction studies the power to discriminate between random variation and treatment effect is poor and the presence or absence of a dosage-related trend can be a critical means of determining the probability of a treatment effect. It has to be kept in mind that in these studies dose responses may be steep, and wide intervals between doses would be inadvisable. If an analysis of dose-response relationships for the effects observed is attempted in a single study, it is recommended to use at least three dose levels and appropriate control groups. If in doubt, a fourth

dose group should be added to avoid excessive dosage intervals. Such a strategy should provide a "no observed adverse effect level" for reproductive aspects. If not, the implication is that the test substance merits a greater depth of investigation and further studies.

9.   Exposure by Different Routes of Administration [III.B. (3.2)]

If it can be shown that one route provides a greater body burden, e.g., area under the curve (AUC), there seems little reason to investigate routes that would provide a lesser body burden or which present severe practical difficulties (e.g. inhalation). Before designing new studies for a new route of administration, existing data on kinetics should be used to determine the necessity of another study.

10.   Kinetics in Pregnant Animals [III.C. (3.3)]

Kinetic investigations in pregnant and lactating animals may pose some problems due to the rapid changes in physiology. It is best to consider this as a two- or three-phase approach. In planning studies kinetic data (often from nonpregnant animals) provide information on the general suitability of the species, and can assist in deciding study designs and choice of dosage. During a study kinetic investigations can provide assurance of accurate dosing or indicate marked deviations from expected patterns.

11.   Examples for Choosing Other Options [IV. (4)]

For compounds causing no lethality at 2 g/kg and no evidence of repeated dose toxicity at 1 g/kg, conduct of a single two-generation study with one control and two test groups (0.5 and 1.0 g/kg) would seem sufficient. However, it might pose the question as to whether the correct species had been chosen or whether the compound was an effective medicine.

For compounds that may be given as a single dose, once in a lifetime (e.g., diagnostics, medicines used in operations), it may be impossible to administer repeated dosages more than twice the human therapeutic dosage for any length of time. A reduced period of treatment allowing a higher dose would seem more appropriate. For females, considerations of human exposure suggest little or no need for exposures beyond the embryonic period.

For dopamine agonists or compounds reducing circulating prolactin levels, female rats are poor models; the rabbit would probably make a better choice

A-6

for all the reproductive toxicity studies, but it does not appear to have been attempted. This also applies to other types of compound when the rabbit shows a pattern of metabolism considerably closer to humans than the rat.

For drugs where alterations in plasma kinetics are seen following repeated administration, the potential for adverse effects on embryo-fetal development may not be fully evaluated in studies according to IV.A.3. (4.1.3). In such cases it may be desirable to extend the period of drug administration to females in a IV.A.1. (4.1.1) study to day 17. With sacrifice at term, both fertility and embryo-fetal development can be assessed.

12.   Premating Treatment [IV.A.1. (4.1.1)]

The design of the fertility study, especially the reduction in the premating period for males, is based on evidence accumulated and reappraisal of the basic research on the process of spermatogenesis that originally prompted the demand for a prolonged premating treatment period. Compounds inducing selective effects on male reproduction are rare; mating with females is an insensitive means of detecting effects on spermatogenesis; good pathological and histopathological examination (e.g., by employing Bouin's fixation, paraffine embedding, transverse sections of 2 to 4 microns for testes, longitudinal sections for epididymides, PAS, and haematoxylin staining) of the male reproductive organs provides a more sensitive and quicker means of detecting effects on spermatogenesis; compounds affecting spermatogenesis almost invariably affect postmeiotic stages; there is no conclusive example of a male reproductive toxicant the effects of which could be detected only by dosing males for 9 to 10 weeks and mating them with females.

Information on potential effects on spermatogenesis can be derived from repeated dose toxicity studies. This allows the investigations in the fertility study to be concentrated on other, more immediate, causes of effect. It is noted that the full sequence of spermatogenesis (including sperm maturation) in rats lasts 63 days. When the available evidence, or lack of it, suggests that the scope of investigations in the fertility study should be increased, or extended from detection to characterization, appropriate studies should be designed to further characterize the effects.

13.   Number of Animals [IV.A.1. (4.1.1), IV.A.2. (4.1.2), IV.A.3. (4.1.3)]

There is very little scientific basis underlying specified group sizes in past

and existing guidelines nor in this one. The numbers specified are educated guesses governed by the maximum study size that can be managed without undue loss of overall study control. This is indicated by the fact that the more expensive the animal is to obtain or keep, the smaller the group size proposed. Ideally, at least the same group size should be required for all species and there is a case for using larger group sizes for less frequently used species such as primates.

It should also be made clear that the numbers required depend on whether or not the group is expected to demonstrate an effect. For a high frequency effect few animals are required, to presume the absence of an effect the number required varies according to the variable (endpoint) being considered, its prevalence in control populations (rare or categorical events), or dispersion around the central tendency (continuous or semicontinuous variables). See also Note 23.

For all but the rarest events (such as malformations, abortions, total litter loss), evaluation of between 16 to 20 litters for litters for rodents and rabbits tends to provide a degree of consistency between studies. Below 16 litters per evaluation, between study results become inconsistent, above 20 to 24 litters per group, consistency and precision are not greatly enhanced. These numbers relate to evaluation. If groups are subdivided for different evaluations the number of animals starting the study should be doubled. Similarly, in studies with 2 breeding generations, 16 to 20 litters would be required for the final evaluation of the litters of the F1 generation. To allow for natural wastage, the starting group size of the F0 generation must be larger.

14.   Mating [IV.A.I. (4.1.1)]

Mating ratios: When both the sexes are being dosed or are of equal consideration in separate male and female studies, the preferred mating ratio is 1:1 because this is the safest option in respect of obtaining good pregnancy rates and avoiding incorrect analysis and interpretation of results.

Mating period and practices: Most laboratories would use a mating period of between 2 and 3 weeks, some remove females as soon as a positive vaginal smear or plug is observed whilst others leave the pairs together. Most rats will mate within the first 5 days of cohabitation (i.e., at the first available estrus), but in some cases females may become pseudopregnant. Leaving the female with the male for about 20 days allows these females to

restart estrus cycles and become pregnant.

15.   Terminal Sacrifice [IV.A.1. (4.1.1)]

Females: When exposure of the females ceases at implantation, termination of females between days 13 and 15 of pregnancy in general is adequate to assess effects on fertility or reproductive function, e.g., to differentiate between implantation and resorption sites.

In general, for detection of adverse effects, it is not thought necessary, in a fertility study, to sacrifice females at day 20/21 of pregnancy in order to gain information on late embryo loss, fetal death, and structural abnormalities.

Males: It would be advisable to delay sacrifice of the males until the outcome of mating is known. In the event of an equivocal result, males could be mated with untreated females to ascertain their fertility or infertility. The males treated as part of study IV.A.1. may also be used for evaluation of toxicity to the male reproductive system if dosing is continued beyond mating and sacrifice delayed.

16.   Observations [IV.A.1. (4.1.1), IV.A.2. (4.1.2), IV.A.3. (4.1.3)]

Daily weighing of pregnant females during treatment can provide useful information. Weighing an animal more frequently than twice weekly during periods other than pregnancy (premating, mating, lactation) may also be advisable for some compounds.

For apparently nonpregnant rats or mice (but not rabbits), ammonium sulphide staining of the uterus might be useful to identify peri-implantation death of embryos.

17.   Treatment of Offspring [IV.A.2. (4.1.2)]

Consequent to derivation from existing guidelines for medicines, this guideline does not fully cover exposures from weaning through puberty, nor does it deal with the possibility of reduced reproductive life span.

To detect adverse effects for medicinal products that may be used in infants and juveniles, special studies (case-by-case designs) involving direct treatment of offspring, at ages to be specified, should be considered.

A-9

18.   Separate Embryotoxicity and Peripostnatal Studies [IV.A.2. (4.1.2)]

        If a prenatal and postnatal study is separated into two studies, one covering the embryonic period the other the fetal period, parturition, and lactation, postnatal evaluation of offspring is required in both studies.

19.   F1-Animals [IV.A.2. (4.1.2)]

        The guideline suggests selection of one male and one female per litter on the evidence that it is feasible to conduct behavioral and other functional tests on the same F1 individuals that will be used for assessment of reproductive function. This has the advantage of allowing cross referencing of performance in different tests at the individual level. It is recognized, however, that some laboratories prefer to select separate sets of animals for behavior testing and for assessment of reproductive function. Which is the most suitable for an individual laboratory will depend upon the combination of tests used and the resources available.

20.   Reduction of Litter Size [IV.A.2. (4.1.2)]

        The value of culling or not culling for detection of effects on reproduction is still under discussion. Whether or not culling is performed, it should be explained by the investigator.

21.   Physical Development, Sensory Functions, Reflexes, and Behavior [IV.A.2. (4.1.2)]

        The best indicator of physical development is bodyweight. Achievement of preweaning landmarks of development such as pinna unfolding, coat growth, incisor eruption, etc., is highly correlated with pup bodyweight. This weight is better related to postcoital time than postnatal time, at least when significant differences in gestation length occur. Reflexes, surface righting, auditory startle, air righting, and response to light are also dependent on physical development.

        Two postweaning landmarks of development that are advised are vaginal opening of females and cleavage of the balanopreputial gland of males. The latter is associated with increasing testosterone levels whereas testis descent is not. These landmarks indicate the onset of sexual maturity and it is advised that bodyweight be recorded at the time of attainment to determine whether any differences from control are specific or related to general growth.

A-10

Functional tests: To date, functional tests have been directed almost exclusively to behavior. Even though a great deal of effort has been expended in this direction it is not possible to recommend specific test methods. Investigators are encouraged to find methods that will assess sensory functions, motor activity, learning, and memory.

22.   Individual Identification and Evaluation of Fetuses [IV.A.3. (4.1.3)]

It must be possible to relate all findings by different techniques (i.e., body weight, external inspection, visceral, and/or skeletal examinations) to single specimen in order to detect patterns of abnormalities. The examination of mid- and low-dose fetuses for visceral and/or skeletal abnormalities may not be necessary where the evaluation of the high-dose and the control groups did not reveal any relevant differences. It is advisable, however, to store the fixed specimen for possible later examination. If fresh dissection techniques are normally used, difficulties with later comparisons involving fixed fetuses should be anticipated.

23.   Inferential Statistics [V (5)]

"Significance" tests (inferential statistics) can be used only as a support for the interpretation of results. The interpretation itself is to be based on biological plausibility. It is unwise to assume that a difference from control values is not biologically relevant simply because it is not "statistically significant." To a lesser extent it can be unwise to assume that a "statistically significant" difference must be biologically relevant. Particularly for low frequency events (e.g., embryonic death, malformations) with one-sided distributions, the statistical power of studies is low. Confidence intervals for relevant quantities can indicate the likely size of the effect. When using statistical procedures, experimental units of comparison should be considered: the litter, not the individual conceptus, the mating pair, when both sexes are treated, the mating pair of the parent generation in a two-generation study.